McKinsey&Company

MCKINSEY GLOBAL INSTITUTE

THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD

DECEMBER 2016

IN COLLABORATION WITH
MCKINSEY ANALYTICS

EXECUTIVE SUMMARY

MCKINSEY GLOBAL INSTITUTE

CELEBRATING 25 YEARS OF INSIGHT

MCKINSEY GLOBAL INSTITUTE

In the 25 years since its founding, the McKinsey Global Institute (MGI) has sought to develop a deeper understanding of the evolving global economy. As the business and economics research arm of McKinsey & Company, MGI aims to provide leaders in the commercial, public, and social sectors with the facts and insights on which to base management and policy decisions. The Lauder Institute at the University of Pennsylvania ranked MGI the world's number-one private-sector think tank in its 2015 Global Think Tank Index.

MGI research combines the disciplines of economics and management, employing the analytical tools of economics with the insights of business leaders. Our "micro-to-macro" methodology examines microeconomic industry trends to better understand the broad macroeconomic forces affecting business strategy and public policy. MGI's in-depth reports have covered more than 20 countries and 30 industries. Current research focuses on six themes: productivity and growth, natural resources, labor markets, the evolution of global financial markets, the economic impact of technology and innovation, and urbanization.

Recent reports have assessed the economic benefits of tackling gender inequality, a new era of global competition, Chinese innovation, and digital globalization. MGI is led by four McKinsey & Company senior partners: Jacques Bughin, James Manyika, Jonathan Woetzel, and Frank Mattern, MGI's chairman. Michael Chui, Susan Lund, Anu Madgavkar, and Jaana Remes serve as MGI partners. Project teams are led by the MGI partners and a group of senior fellows, and include consultants from McKinsey offices around the world. These teams draw on McKinsey's global network of partners and industry and management experts. Input is provided by the MGI Council, which co-leads projects and provides guidance; members are Andres Cadena, Richard Dobbs, Katy George, Rajat Gupta, Eric Hazan, Eric Labaye, Acha Leke, Scott Nyquist, Gary Pinkus, Shirish Sankhe, Oliver Tonby, and Eckart Windhagen. In addition, leading economists, including Nobel laureates, act as research advisers.

The partners of McKinsey fund MGI's research; it is not commissioned by any business, government, or other institution. For further information about MGI and to download reports, please visit **www.mckinsey.com/mgi**.

MCKINSEY ANALYTICS

McKinsey Analytics helps clients achieve better performance through data, working together with them to build analytics-driven organizations and providing end-to-end support covering strategy, operations, data science, implementation, and change management. Engagements range from use-case specific applications to full-scale analytics transformations. Teams of McKinsey consultants, data scientists, and engineers work with clients to identify opportunities, assess available data, define solutions, establish optimal hosting environments, ingest data, develop cutting-edge algorithms, visualize outputs, and assess impact while building capabilities to sustain and expand it.

Copyright © McKinsey & Company 2016

THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD

DECEMBER 2016

Nicolaus Henke | London
Jacques Bughin | Brussels
Michael Chui | San Francisco
James Manyika | San Francisco
Tamim Saleh | London
Bill Wiseman | Taipei
Guru Sethupathy | Washington, DC

IN BRIEF

THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD

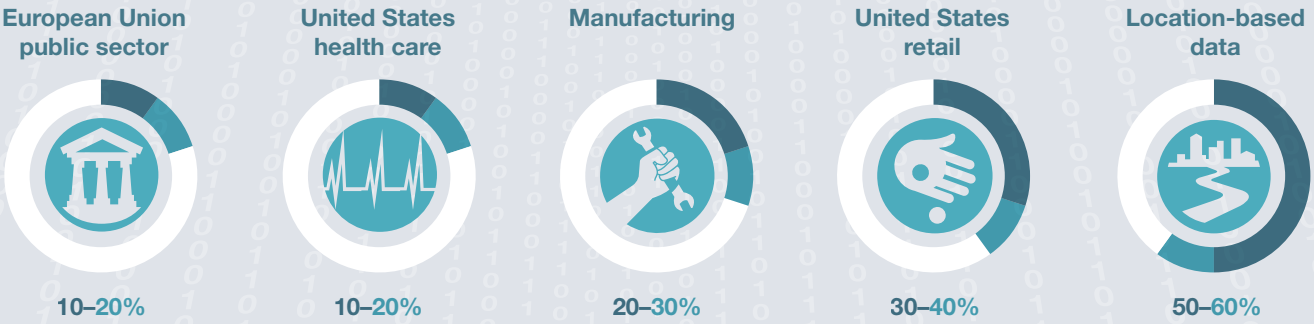
Data and analytics capabilities have made a leap forward in recent years. The volume of available data has grown exponentially, more sophisticated algorithms have been developed, and computational power and storage have steadily improved. The convergence of these trends is fueling rapid technology advances and business disruptions.

- Most companies are capturing only a fraction of the potential value from data and analytics. Our 2011 report estimated this potential in five domains; revisiting them today shows a great deal of value still on the table. The greatest progress has occurred in location-based services and in retail, both areas with digital native competitors. In contrast, manufacturing, the public sector, and health care have captured less than 30 percent of the potential value we highlighted five years ago. Further, new opportunities have arisen since 2011, making the gap between the leaders and laggards even bigger.
- The biggest barriers companies face in extracting value from data and analytics are organizational; many struggle to incorporate data-driven insights into day-to-day business processes. Another challenge is attracting and retaining the right talent—not only data scientists but business translators who combine data savvy with industry and functional expertise.
- Data and analytics are changing the basis of competition. Leading companies are using their capabilities not only to improve their core operations but to launch entirely new business models. The network effects of digital platforms are creating a winner-take-most dynamic in some markets.
- Data is now a critical corporate asset. It comes from the web, billions of phones, sensors, payment systems, cameras, and a huge array of other sources—and its value is tied to its ultimate use. While data itself will become increasingly commoditized, value is likely to accrue to the owners of scarce data, to players that aggregate data in unique ways, and especially to providers of valuable analytics.
- Data and analytics underpin several disruptive models. Introducing new types of data sets ("orthogonal data") can disrupt industries, and massive data integration capabilities can break through organizational and technological silos, enabling new insights and models. Hyperscale digital platforms can match buyers and sellers in real time, transforming inefficient markets. Granular data can be used to personalize products and services—and, most intriguingly, health care. New analytical techniques can fuel discovery and innovation. Above all, data and analytics can enable faster and more evidence-based decision making.
- Recent advances in machine learning can be used to solve a tremendous variety of problems—and deep learning is pushing the boundaries even further. Systems enabled by machine learning can provide customer service, manage logistics, analyze medical records, or even write news stories. The value potential is everywhere, even in industries that have been slow to digitize. These technologies could generate productivity gains and an improved quality of life—along with job losses and other disruptions. Previous MGI research found that 45 percent of work activities could potentially be automated by currently demonstrated technologies; machine learning can be an enabling technology for the automation of 80 percent of those activities. Breakthroughs in natural language processing could expand that impact even further.

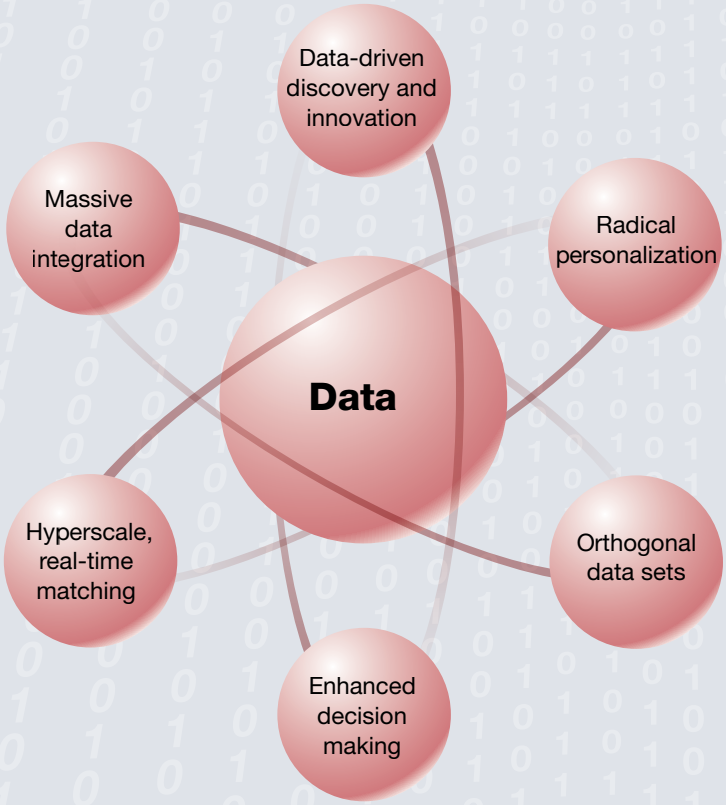
Data and analytics are already shaking up multiple industries, and the effects will only become more pronounced as adoption reaches critical mass. An even bigger wave of change is looming on the horizon as deep learning reaches maturity, giving machines unprecedented capabilities to think, problem-solve, and understand language. Organizations that are able to harness these capabilities effectively will be able to create significant value and differentiate themselves, while others will find themselves increasingly at a disadvantage.

The age of analytics:
Competing in a data-driven world

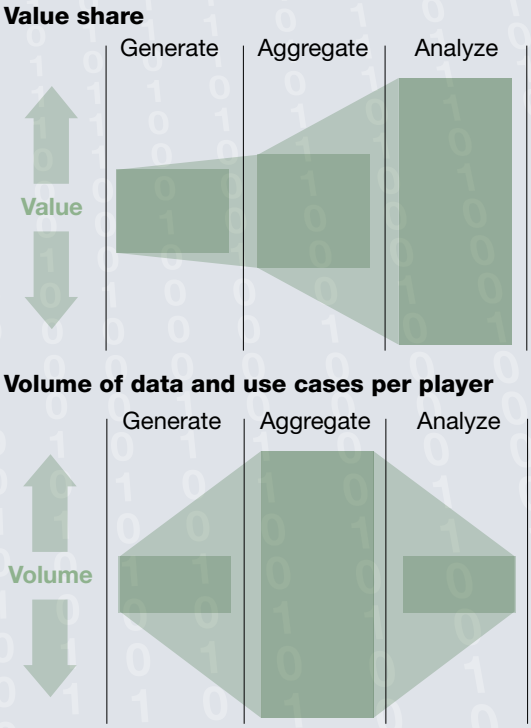
Only a fraction of the value we envisioned in 2011 has been captured to date



Data and analytics fuel 6 disruptive models that change the nature of competition

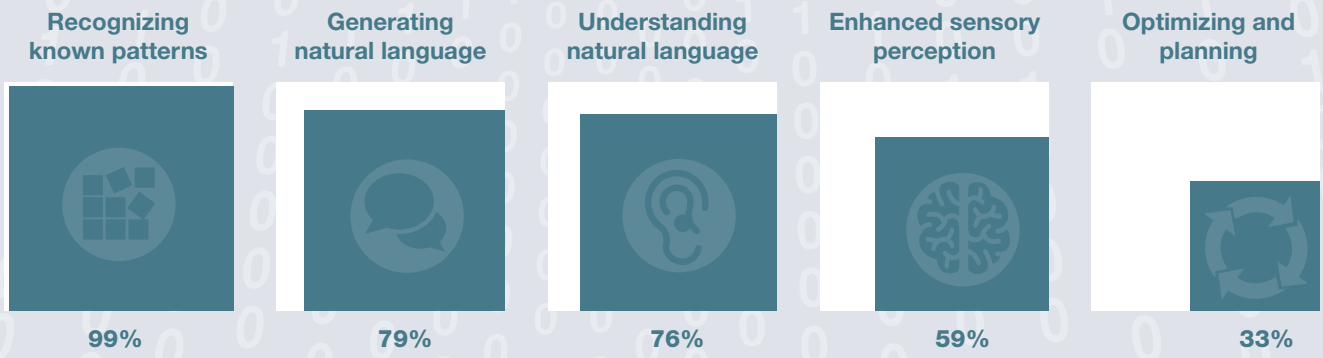


As data ecosystems evolve, value will accrue to providers of analytics, but some data generators and aggregators will have unique value



Machine learning has broad applicability in many common work activities

Percent of work activities that require:



MCKINSEY GLOBAL INSTITUTE

McKinsey&Company

© B. Busco/Getty Images

EXECUTIVE SUMMARY

Back in 2011, the McKinsey Global Institute published a report highlighting the transformational potential of big data.[1] Five years later, we remain convinced that this potential has not been overhyped. In fact, we now believe that our 2011 analyses gave only a partial view. The range of applications and opportunities has grown even larger today.

The convergence of several technology trends is accelerating progress. The volume of data continues to double every three years as information pours in from digital platforms, wireless sensors, and billions of mobile phones. Data storage capacity has increased, while its cost has plummeted. Data scientists now have unprecedented computing power at their disposal, and they are devising ever more sophisticated algorithms.

The companies at the forefront of these trends are using their capabilities to tackle business problems with a whole new mindset. In some cases, they have introduced data-driven business models that have taken entire industries by surprise. Digital natives have an enormous advantage, and to keep up with them, incumbents need to apply data and analytics to the fundamentals of their existing business while simultaneously shifting the basis of competition. In an environment of increasing volatility, legacy organizations need to have one eye on high-risk, high-reward moves of their own, whether that means entering new markets or changing their business models. At the same time, they have to apply analytics to improve their core operations. This may involve identifying new opportunities on the revenue side, using analytics insights to streamline internal processes, and building mechanisms for experimentation to enable continuous learning and feedback.

Organizations that pursue this two-part strategy will be ready to take advantage of opportunities and thwart potential disruptors—and they have to assume that those disruptors are right around the corner. Data and analytics have altered the dynamics in many industries, and change will only accelerate as machine learning and deep learning develop capabilities to think, problem-solve, and understand language. The potential uses of these technologies are remarkably broad, even for sectors that have been slow to digitize. As we enter a world of self-driving cars, personalized medicine, and intelligent robots, there will be enormous new opportunities as well as significant risks—not only for individual companies but for society as a whole.

MOST COMPANIES ARE CAPTURING ONLY A FRACTION OF THE POTENTIAL VALUE OF DATA AND ANALYTICS

Turning a world full of data into a data-driven world is an idea that many companies have found difficult to pull off in practice. Our 2011 report estimated the potential for big data and analytics to create value in five specific domains. Revisiting them today shows both uneven progress and a great deal of that value still on the table (Exhibit E1).

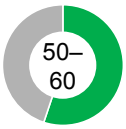
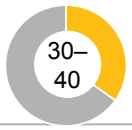
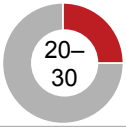
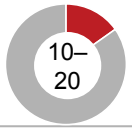
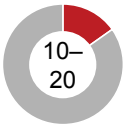
We see the greatest progress in location-based services and in US retail. In contrast, adoption is lagging in manufacturing, the EU public sector, and US health care. Incentive problems and regulatory issues pose additional barriers to adoption in the public sector and health care. In several cases, incumbent stakeholders that would have the most to lose from the kinds of changes data and analytics could enable also have a strong influence on regulations, a factor that could hinder adoption.

---

[1] *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, June 2011.

Exhibit E1

There has been uneven progress in capturing value from data and analytics

	Potential impact: 2011 research	Value captured %	Major barriers
Location-based data	▪ $100 billion+ revenues for service providers ▪ Up to $700 billion value to end users		▪ Penetration of GPS-enabled smartphones globally
US retail[1]	▪ 60%+ increase in net margin ▪ 0.5–1.0% annual productivity growth		▪ Lack of analytical talent ▪ Siloed data within companies
Manufacturing[2]	▪ Up to 50% lower product development cost ▪ Up to 25% lower operating cost ▪ Up to 30% gross margin increase		▪ Siloed data in legacy IT systems ▪ Leadership skeptical of impact
EU public sector[3]	▪ ~€250 billion value per year ▪ ~0.5% annual productivity growth		▪ Lack of analytical talent ▪ Siloed data within different agencies
US health care	▪ $300 billion value per year ▪ ~0.7% annual productivity growth		▪ Need to demonstrate clinical utility to gain acceptance ▪ Interoperability and data sharing

1 Similar observations hold true for the EU retail sector.
2 Manufacturing levers divided by functional application.
3 Similar observations hold true for other high-income country governments.

SOURCE: Expert interviews; McKinsey Global Institute analysis

- **Location-based services:** GPS-enabled smartphones have put mapping technology in the pockets of billions of users. The markets for global positioning system navigation devices and services, mobile phone location-based service applications, and geo-targeted mobile advertising services have reached 50 to 60 percent of the value we envisioned in 2011. End consumers are capturing the lion's share of the benefits, mostly through time and fuel savings as well as new types of mobile services. Beyond the value we envisioned in 2011, there are growing opportunities for businesses to use geospatial data to track assets, teams, and customers across dispersed locations in order to generate new insights and improve efficiency.

- **US retail:** Retailers can mine a trove of transaction-based and behavioral data from their customers. Thin margins (especially in the grocery sector) and pressure from industry-leading early adopters such as Amazon and Walmart have created strong incentives to put that data to work in everything from cross-selling additional products to reducing costs throughout the entire value chain. The US retail sector has realized 30 to 40 percent of the potential margin improvements and productivity growth we envisioned in 2011, but again, a great deal of value has gone to consumers.

- **Manufacturing:** Manufacturing industries have achieved only about 20 to 30 percent of the potential value we estimated in 2011—and most has gone to a handful of industry leaders. Within research and design, design-to-value applications have seen the greatest uptick in adoption, particularly among carmakers. Some industry leaders have developed digital models of the entire production process ("digital factories"). More companies have integrated sensor data-driven operations analytics, often reducing

operating costs by 5 to 15 percent. After-sales servicing offers are beginning to be based on real-time surveillance and predictive maintenance.

10-20%
of the potential value has been captured in the public sector and health care

- **The EU public sector:** Our 2011 report analyzed how the European Union's public sector could use data and analytics to make government services more efficient, reduce fraud and errors in transfer payments, and improve tax collection, potentially achieving some €250 billion worth of annual savings. But only about 10 to 20 percent of this has materialized. Some agencies have moved more interactions online, and many (particularly tax agencies) have introduced pre-filled forms. But across Europe and other advanced economies, adoption and capabilities vary greatly. The complexity of existing systems and the difficulty of attracting scarce analytics talent with public-sector salaries have slowed progress. Despite this, we see even wider potential today for societies to use analytics to make more evidence-based decisions in many aspects of government.

- **US health care:** To date, only 10 to 20 percent of the opportunities we outlined in 2011 have been realized by the US health-care sector. A range of barriers—including a lack of incentives, the difficulty of process and organizational changes, a shortage of technical talent, data-sharing challenges, and regulations—have combined to limit adoption. Within clinical operations, the biggest success has been the shift to electronic medical records, although the vast stores of data they contain have not yet been fully mined. While payers have been slow to capitalize on big data for accounting and pricing, a growing industry now aggregates and synthesizes clinical records, and analytics have taken on new importance in public health surveillance. Many pharmaceutical firms are using analytics in R&D, particularly in streamlining clinical trials. While the health-care sector continues to lag in adoption, there are enormous unrealized opportunities to transform clinical care and deliver personalized medicine (a topic we will return to below).

LEGACY COMPANIES HAVE TO OVERCOME HURDLES TO ACCELERATE THEIR ANALYTICS TRANSFORMATION

The relatively slow pace of progress in some of the domains described above points to the fact that many companies that have begun to deploy data and analytics have not realized the full value. Some have responded to competitive pressure by making large technology investments but have failed to make the organizational changes needed to make the most of them.

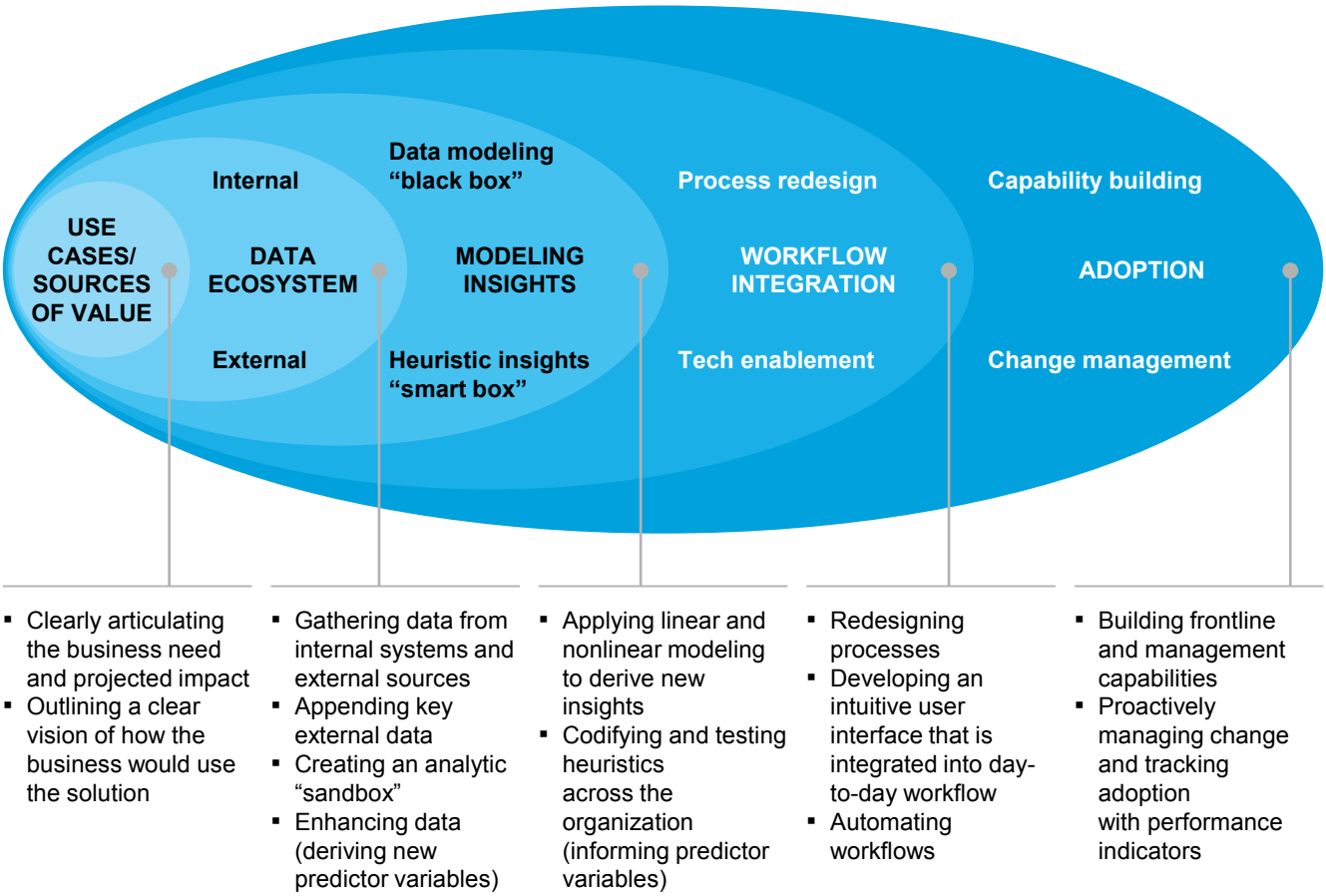
An effective transformation strategy can be broken down into several components (Exhibit E2). The first step should be asking some fundamental questions to shape the strategic vision: What will data and analytics be used for? How will the insights drive value? How will the value be measured? The second element is building out the underlying data architecture as well as data collection or generation capabilities. Many incumbents struggle with switching from legacy data systems to a more nimble and flexible architecture to store and harness big data. They may also need to digitize their operations more fully in order to capture more data from their customer interactions, supply chains, equipment, and internal processes. Looking at a wide variety of indicators that measure digitization, we see a striking gap between leading firms and average firms on this front.[2] The third piece is acquiring the analytics capabilities needed to derive insights from data; organizations may choose to add in-house capabilities or outsource to specialists. The fourth component is a common stumbling block: changing business processes to incorporate data insights into the actual workflow. This requires getting the right data insights into the hands of the right personnel. Finally, organizations need to build the capabilities of executives and mid-level managers to understand how to use data-driven insights—and to begin to rely on them as the basis for making decisions.

[2] *Digital America: A tale of the haves and have-mores*, McKinsey Global Institute, December 2015; and *Digital Europe: Pushing the frontier, capturing the benefits*, McKinsey Global Institute, June 2016.

Where digital native companies were built for analytics, legacy companies have to do the hard work of overhauling or changing existing systems. Neglecting any of these elements can limit the potential value of analytics or even leave an organization vulnerable to being disrupted. It may be a difficult transition, but some long-established names, including GE and Union Pacific, have managed to pull it off.

Exhibit E2

Successful data and analytics transformation requires focusing on five elements


USE CASES/ SOURCES OF VALUE
DATA ECOSYSTEM
Internal
External
MODELING INSIGHTS
Data modeling "black box"
Heuristic insights "smart box"
WORKFLOW INTEGRATION
Process redesign
Tech enablement
ADOPTION
Capability building
Change management
▪ Clearly articulating the business need and projected impact
▪ Outlining a clear vision of how the business would use the solution
▪ Gathering data from internal systems and external sources
▪ Appending key external data
▪ Creating an analytic "sandbox"
▪ Enhancing data (deriving new predictor variables)
▪ Applying linear and nonlinear modeling to derive new insights
▪ Codifying and testing heuristics across the organization (informing predictor variables)
▪ Redesigning processes
▪ Developing an intuitive user interface that is integrated into day-to-day workflow
▪ Automating workflows
▪ Building frontline and management capabilities
▪ Proactively managing change and tracking adoption with performance indicators

SOURCE: McKinsey Analytics; McKinsey Global Institute analysis

THERE IS A CONTINUING SHORTAGE OF ANALYTICS TALENT

Across the board, companies report that finding the right talent is the biggest hurdle they face in trying to integrate data and analytics into their existing operations. In a recent McKinsey & Company survey, approximately half of executives across geographies and industries reported greater difficulty recruiting analytical talent than filling any other kind of role. Forty percent say retention is also an issue.[3]

Data scientists, in particular, are in high demand. Our 2011 report hypothesized that demand for data scientists would outstrip supply. This is in fact what we see in the labor market today, despite the fact that universities are adding data and analytics programs and that other types of training programs are proliferating. Average wages for data scientists in the United States rose by approximately 16 percent a year from 2012 to 2014.[4] This far

[3] "The need to lead in data and analytics," McKinsey & Company survey, McKinsey.com, April 2016, available at http://www.mckinsey.com/business-functions/business-technology/our-insights/the-need-to-lead-in-data-and-analytics. The online survey, conducted in September 2015, garnered responses from more than 500 executives across a variety of regions, industries, and company sizes.

[4] *Beyond the talent shortage: How tech candidates search for jobs*, Indeed.com, September 2015.

outstrips the less than 2 percent increase in nominal average wages across all occupations in US Bureau of Labor Statistics data. The scarcity of elite data scientists has even been a factor in some acquisitions of cutting-edge artificial intelligence (AI) startups; deals can command around $5 million to $10 million per employee.

This trend is likely to continue in the near term. While we estimate that the number of graduates from data science programs could increase by a robust 7 percent per year, our high-case scenario projects even greater (12 percent) annual growth in demand, which would lead to a shortfall of some 250,000 data scientists. But a countervailing force could ease this imbalance in the medium term: data preparation, which accounts for more than 50 percent of data science work, could be automated. Whether that dampens the demand for data scientists or simply enables data scientists to shift their work toward analysis and other activities remains to be seen.

Many organizations focus on the need for data scientists, assuming their presence alone will enable an analytics transformation. But another equally vital role is that of the business translator who serves as the link between analytical talent and practical applications to business questions. In addition to being data savvy, business translators need to have deep organizational knowledge and industry or functional expertise. This enables them to ask the data science team the right questions and to derive the right insights from their findings. It may be possible to outsource analytics activities, but business translator roles require proprietary knowledge and should be more deeply embedded into the organization. Many organizations are building these capabilities from within.

2M–4M
projected US demand for business translators over the next decade

We estimate there could be demand for approximately two million to four million business translators in the United States alone over the next decade. Given the roughly 9.5 million US graduates in business and in the STEM fields of science, technology, engineering, and mathematics expected over the same period, nearly 20 to 40 percent of these graduates would need to go into business translator roles to meet demand.[5] Today that figure is only about 10 percent. To reduce this mismatch, wages may have to increase, or more companies will need to implement their own training programs.[6]

As data grows more complex, distilling it and bringing it to life through visualization is becoming critical to help make the results of data analyses digestible for decision makers. We estimate that demand for visualization grew roughly 50 percent annually from 2010 to 2015.[7] In many instances today, organizations are seeking data scientist or business translator candidates who can also execute visualizations. However, we expect that medium-size and large organizations, as well as analytics service providers, will increasingly create specialized positions for candidates who combine a strong understanding of data with user interface, user experience, and graphic design skills.

ANALYTICS LEADERS ARE CHANGING THE NATURE OF COMPETITION AND CONSOLIDATING BIG ADVANTAGES

There are now major disparities in performance between a small group of technology leaders and the average company—in some cases creating winner-take-most dynamics. Leaders such as Apple, Alphabet/Google, Amazon, Facebook, Microsoft, GE, and Alibaba Group have established themselves as some of the most valuable companies in the world.[8] The same trend can be seen among privately held companies. The leading global “unicorns”

5 Non-STEM graduates with quantitative skills can also fill business translator roles.

6 Sam Ransbotham, David Kiron, and Pamela Kirk Prentice, “The talent dividend: Analytics talent is driving competitive advantage at data-oriented companies,” *MIT Sloan Management Review*, April 25, 2015.

7 Based on using the Burning Glass job postings database to search for postings including any of the following skills: data visualization, Tableau, Qlikview, and Spotfire. Normalized with the total number of job postings.

8 Michael Chui and James Manyika, “Competition at the digital edge: ‘Hyperscale businesses,’” *McKinsey Quarterly*, March 2015.

tend to be companies with business models predicated on data and analytics, such as Uber, Lyft, Didi Chuxing, Palantir, Flipkart, Airbnb, DJI, Snapchat, Pinterest, BlaBlaCar, and Spotify. These companies differentiate themselves through their data and analytics assets, processes, and strategies.

The relative value of various assets has shifted. Where previous titans of industry poured billions into factories and equipment, the new leaders invest heavily in digital platforms, data, and analytical talent. New digital native players can circumvent traditional barriers to entry, such as the need to build traditional fixed assets, which enables them to enter markets with surprising speed. Amazon challenged the rest of the retail sector without building stores (though it does have a highly digitized physical distribution network), "fintechs" are providing financial services without physical bank branches, Netflix is changing the media landscape without connecting cables to customers' homes, and Airbnb has introduced a radical new model in the hospitality sector without building hotels. But some digital natives are now erecting new barriers to entry themselves; platforms may have such strong network effects that they give operators a formidable advantage within a given market.

The leading firms have a remarkable depth of analytical talent deployed on a variety of problems—and they are actively looking for ways to enter other industries. These companies can take advantage of their scale and data insights to add new business lines, and those expansions are increasingly blurring traditional sector boundaries.[9] Apple and Alibaba, for instance, have introduced financial products and services, while Google is developing autonomous cars. The importance of data has also upended the traditional relationship between organizations and their customers since every interaction generates information. Sometimes the data itself is so prized that companies offer free services in order to obtain it; this is the case with Facebook, LinkedIn, Pinterest, Twitter, Tencent, and many others. An underlying barter system is at work, particularly in the consumer space, as individuals gain access to digital services in return for data about their behaviors and transactions.

THE VALUE OF DATA DEPENDS ON ITS ULTIMATE USE, AND ECOSYSTEMS ARE EVOLVING TO HELP COMPANIES CAPTURE THAT VALUE

Data is at the heart of the disruptions occurring across the economy. It has become a critical corporate asset, and business leaders want to know what the information they hold is worth. But its value is tied to how it will be used and by whom. A piece of data may yield nothing, or it may yield the key to launching a new product line or cracking a scientific question. It might affect only a small percentage of a company's revenue today, but it could be a driver of growth in the future.

Not all data are created equal

Part of the challenge in valuing data is its sheer diversity. Some of the broad categories include behavioral data (capturing actions in both digital and physical environments), transactional data (records of business dealings), ambient or environmental data (conditions in the physical world monitored and captured by sensors), geospatial data, reference material or knowledge (news stories, textbooks, reference works, literature, and the like), and public records. Some data are structured (that it, easily expressed in rows and columns), while images, audio, and video are unstructured. Data can also come from the web, social media, industrial sensors, payment systems, cameras, wearable devices, and human entry. Billions of mobile phones, in particular, are capturing images, video, and location data. On the demand side, data can provide insights for diverse uses, some of which are more valuable than others.

[9] *Playing to win: The new global competition for corporate profits*, McKinsey Global Institute, September 2015.

Over the long term, value will likely accrue to providers of analytics and data platform owners

Many organizations are hungry to use data to grow and improve performance—and multiple players see market opportunities in this explosion of demand. There are typically many steps between raw data and actual usage, and there are openings to add value at various points along the way. To simplify, we focused on three categories of players in the data ecosystem, recognizing that some players might fill more than one role.

- **Data generation and collection:** The source and platform where data are initially captured.
- **Data aggregation:** Processes and platforms for combining data from multiple sources.
- **Data analysis:** The gleaning of insights from data that can be acted upon.

Usually, the biggest opportunities are unlikely to be in directly monetizing data. As data become easier to collect and as storage costs go down, most data are becoming more commoditized. Proxies now exist for data that were once scarce; Google Trends, for instance, offers a free proxy for public sentiment data that previously would have been collected through phone surveys.

However, there are important exceptions to the commoditization trend. When access is limited by physical barriers or collection is expensive, data will hold its value. An important case in which value can accrue to data generation and collection involves market-making or social media platforms with strong network effects. In certain arenas, a small number of players establish such critical mass that they are in a position to collect and own the vast majority of user behavior data generated in these ecosystems. But in the absence of these types of exceptional supply constraints, simply selling raw data is likely to generate diminishing returns over time.

Another role in the data ecosystem involves aggregating information from different sources. In general, this capability is becoming more accessible and less expensive, but this role can be valuable when certain conditions apply. Data aggregation adds value when combining, processing, and aggregating data is technically difficult or organizationally challenging (for example, when aggregating involves coordinating access across diverse sources). Some companies have built business models around serving as third-party aggregators for competitors within a given industry, and this model has the potential to create network effects as well.

The third part of the data ecosystem, analytics, is where we expect to see the biggest opportunities in the future. The provider of analytics understands the value being generated by those insights and is thus best positioned to capture a portion of that value. Data analytics tools, like other software, already command large margins. Combining analytical tools with business insights for decision makers is likely to multiply the value even further. Increasingly complex data and analytics will require sophisticated translation, and use cases will be very firm-specific. Bad analysis can destroy the potential value of high-quality data, while great analysis can squeeze insights from even mediocre data. In addition, the scarcity of analytics talent is driving up the cost of these services. Given the size of the opportunities, firms in other parts of the ecosystem are scrambling to stake out a niche in the analytics market. Data aggregators are offering to integrate clients' data and perform analysis as a service. One-stop shops offering integrated technology stacks are adding analytics capabilities, such as IBM Watson, as are other professional services and business intelligence firms.

SIX DISRUPTIVE DATA-DRIVEN MODELS AND CAPABILITIES ARE RESHAPING SOME INDUSTRIES—AND COULD SOON TRANSFORM MANY MORE

Certain characteristics of a given market (such as inefficient matching, information asymmetries, and human biases and errors) open the door to disruption. They set the stage for six archetypes to have a major effect (Exhibit E3). In each of these models, the introduction of new data is a key enabler.

Exhibit E3

Data and analytics underpin six disruptive models, and certain characteristics make individual domains susceptible

Indicators of potential for disruption:

- Assets are underutilized due to inefficient signaling
- Supply/demand mismatch
- Dependence on large amounts of personalized data
- Data is siloed or fragmented
- Large value in combining data from multiple sources
- R&D is core to the business model
- Decision making is subject to human biases
- Speed of decision making limited by human constraints
- Large value associated with improving accuracy of prediction

Archetype of disruption	Domains that could be disrupted
Business models enabled by orthogonal data	▪ Insurance ▪ Health care ▪ Human capital/talent
Hyperscale, real-time matching	▪ Transportation and logistics ▪ Automotive ▪ Smart cities and infrastructure
Radical personalization	▪ Health care ▪ Retail ▪ Media ▪ Education
Massive data integration capabilities	▪ Banking ▪ Insurance ▪ Public sector ▪ Human capital/talent
Data-driven discovery	▪ Life sciences and pharmaceuticals ▪ Material sciences ▪ Technology
Enhanced decision making	▪ Smart cities ▪ Health care ▪ Insurance ▪ Human capital/talent

SOURCE: McKinsey Global Institute analysis

Bringing in orthogonal data can change the basis of competition

As data proliferate, many new types, from new sources, can be brought to bear on any problem. In industries where most incumbents have become used to relying on a certain kind of standardized data to make decisions, bringing in fresh types of data sets to supplement those already in use can change the basis of competition. New entrants with privileged access to these "orthogonal" data sets can pose a uniquely powerful challenge to incumbents. We see this playing out in property and casualty insurance, where new companies have entered the marketplace with telematics data that provides insight into driving behavior. This is orthogonal to the demographic data that had previously been used for underwriting. Other domains could be fertile ground for bringing in orthogonal data from the internet of things (IoT). Connected light fixtures, which sense the presence of people in a room and have been sold with the promise of reducing energy usage, generate "data exhaust" that property managers can use to optimize physical space planning. Even in human resources, some organizations have secured employee buy-in to wear devices that capture data and yield insights into the "real" social networks that exist in the workplace, enabling these organizations to optimize collaboration through changes in work spaces.

Orthogonal data will rarely replace the data that are already in use in a domain; it is more likely that an organization will integrate orthogonal data with existing data. Within the other

archetypes below are several examples of orthogonal data being combined with existing data to create new business models and improve performance.

Hyperscale platforms can match supply and demand in real time

Digital platforms provide marketplaces that connect sellers and buyers for many products and services. Some platform operators using data and analytics to do this in real time and on an unprecedented scale—and this can be transformative in markets where supply and demand matching has been inefficient.

In personal transportation, ride-sharing services use geospatial mapping technology to collect crucial data about the precise location of passengers and available drivers in real time. The introduction of this new type of data enabled efficient and instant matching, a crucial innovation in this market. In addition, the data can be analyzed at the aggregate level for dynamic pricing adjustments to help supply and demand adjust. The typical personally owned car is estimated to sit idle 85 to 95 percent of the time, making it a hugely underutilized asset. Platforms such as Uber, Lyft, and Chinese ride-sharing giant Didi Chuxing have been able to expand rapidly without acquiring huge fleets themselves, making it easy for new drivers to put their own underutilized assets to work.

Up to
$2.5T
potential economic impact from continued adoption of mobility services by 2025

By 2030 mobility services, such as ride sharing and car sharing, could account for more than 15 to 20 percent of total passenger vehicle miles globally. This growth—and the resulting hit to the taxi industry—may be only a hint of what is to come. Automakers are the biggest question mark. While sales will likely continue to grow in absolute numbers, we estimate that the shift toward mobility services could halve the growth rate of global vehicle sales by 2030. Consumers could save on car purchases, fuel, and parking. If mobility services attain 10 to 30 percent adoption among low-mileage urban vehicle users, the ensuing economic impact could reach $845 billion to some $2.5 trillion globally by 2025. Some of this value will surely go to consumer surplus, while some will go to the providers of these platforms and mobility services.

Data and analytics enable “radical personalization”

Data and analytics can reveal finer levels of distinctions, and one of the most powerful uses is micro-segmenting a population based on the characteristics of individuals. Using the resulting insights to personalize products and services on a wide scale is changing the fundamentals of competition in many sectors, including education, travel and leisure, media, retail, and advertising.

This capability could have profound implications for the way health care is delivered if the sector can incorporate the behavioral, genetic, and molecular data connected with many individual patients. The declining costs of genome sequencing, the advent of proteomics, and the growth of real-time monitoring technologies make it possible to generate this kind of new, ultra-granular data. These data can reshape health care in two profound ways. First, they can help address information asymmetries and incentive problems in the health-care system. Now that a more complete view of the patient is available, incentives could be changed for hospitals and other providers to shift their focus from disease treatment to wellness and prevention, saving huge sums on medical expenditures and improving the quality of life. Second, having more granular and complete data on individual patients can make treatments more precise. Pharmaceutical and medical device companies have enormous possibilities in R&D for accelerating drug discovery, although they will be challenged to create new business models to deliver treatments tailored to smaller, targeted patient populations. Treatments, dosages, and care settings can be personalized to individuals, leading to more effective outcomes with fewer side effects and reduced costs.

Personalized medicine could reduce health-care costs while allowing people to enjoy longer, healthier, and more productive lives. The total impact could range from $2 trillion

to $10 trillion. The wide range depends on the many uncertainties involved, including how rapidly the health-care system can adapt and whether R&D applications produce breakthrough treatments.

Massive data integration capabilities can break down organizational silos

The first step in creating value from data and analytics is accessing all the information that is relevant to a given problem. This may involve generating the data, accessing it from new sources, breaking silos within an organization to link existing data, or all of the above. Combining and integrating large stores of data from all of these varied sources has incredible potential to yield insights, but many organizations have struggled with creating the right structure for that synthesis to take place.

Up to
$260B
potential global impact of massive data integration in retail banking

Retail banking, for instance, is an industry rich with data on customers' transactions, financial status, and demographics. But few institutions have made the most of the data due to internal barriers and the variable quality of the information itself. Surmounting these barriers is critical now that social media, call center discussions, video footage from branches, and data acquired from external sources and partners can be used to form a more complete picture of customers. Massive data integration has significant potential for retail banks. It can enable better cross-selling, the development of personalized products, dynamic pricing, better risk assessment, and more effective marketing—and it can help firms achieve more competitive cost structures than many incumbent institutions. All told, we estimate a potential economic impact of $110 billion to $170 billion in the retail banking industry in developed markets and approximately $60 billion to $90 billion in emerging markets.

Additionally, companies in other sectors can become part of the financial services ecosystem if they bring in orthogonal data—such as non-financial data that provides a more comprehensive and detailed view of the customer. These players may have large customer bases and advanced analytics capabilities created for their core businesses, and they can use these advantages to make rapid moves across sector boundaries. Alibaba's creation of Alipay and Apple's unveiling of Apple Pay are prime examples of this trend.

Data and analytics can fuel discovery and innovation

One of the main components of productivity growth, innovation can be applied to both processes and products. Throughout history, innovative ideas have sprung from human ingenuity and creativity—but now data and algorithms can support, enhance, or even replace human ingenuity in some instances.

In the realm of process innovation, data and analytics are helping organizations determine how to structure teams, resources, and workflows. High-performing teams can be many times more productive than low-performing teams, so understanding this variance and how to build more effective collaboration is a huge opportunity for organizations. This involves looking at issues such as the complementarity of skills, optimal team sizes, whether teams need to work together in person, what past experience or training is important, and even how their personalities may mesh. Data and analytics can test hypotheses and find new patterns that may not have even occurred to managers. Vast amounts of email, calendar, locational, and other data are available to understand how people work together and communicate, all of which can lead to new insights about improving performance.

In product innovation, data and analytics can transform research and development in areas such as materials science, synthetic biology, and life sciences. Leading pharmaceutical companies are using data and analytics to aid with drug discovery. Data from a variety of sources could better determine the chemical compounds that would serves as effective drug treatments for a variety of diseases. AstraZeneca and Human Longevity are partnering

to build a database of one million genomic and health records along with 500,000 DNA samples from clinical trials. The associations and patterns that can be gleaned from that data could prove to be immensely valuable in advancing scientific and drug development breakthroughs.

Algorithms can support and enhance human decision making

When humans make decisions, the process is often muddy, biased, or limited by our inability to process information overload. Data and analytics can change all that by bringing in more data points from new sources, breaking down information asymmetries, and adding automated algorithms to make the process instantaneous. As the sources of data grow richer and more diverse, there are many ways to use the resulting insights to make decisions faster, more accurate, more consistent, and more transparent.

There are many examples of how this can play out in industries and domains across the economy. Smart cities, for example, are one of the most promising settings for applying the ability of machines and algorithms to process huge quantities of information in a fraction of the time it takes humans. Using sensors to improve traffic flows and the internet of things to enable utilities to reduce waste and keep infrastructure systems working at top efficiency are just two of the myriad possible municipal applications. One of the most promising applications of data and analytics is in the prevention of medical errors. Advanced analytical support tools can flag potential allergies or dangerous drug interactions for doctors and pharmacists alike, ensuring that their decisions are consistent and reliable. And finally, perhaps no area of human decision making is quite as opaque and clouded by asymmetric information as hiring. Data and analytics have the potential to create a more transparent labor market by giving employers and job seekers access to data on the supply and demand for particular skills, the wages associated with various jobs, and the value of different degree programs.

THE FRONTIERS OF MACHINE LEARNING, INCLUDING DEEP LEARNING, HAVE RELEVANCE IN EVERY INDUSTRY AND WIDE-RANGING POTENTIAL TO SOLVE PROBLEMS

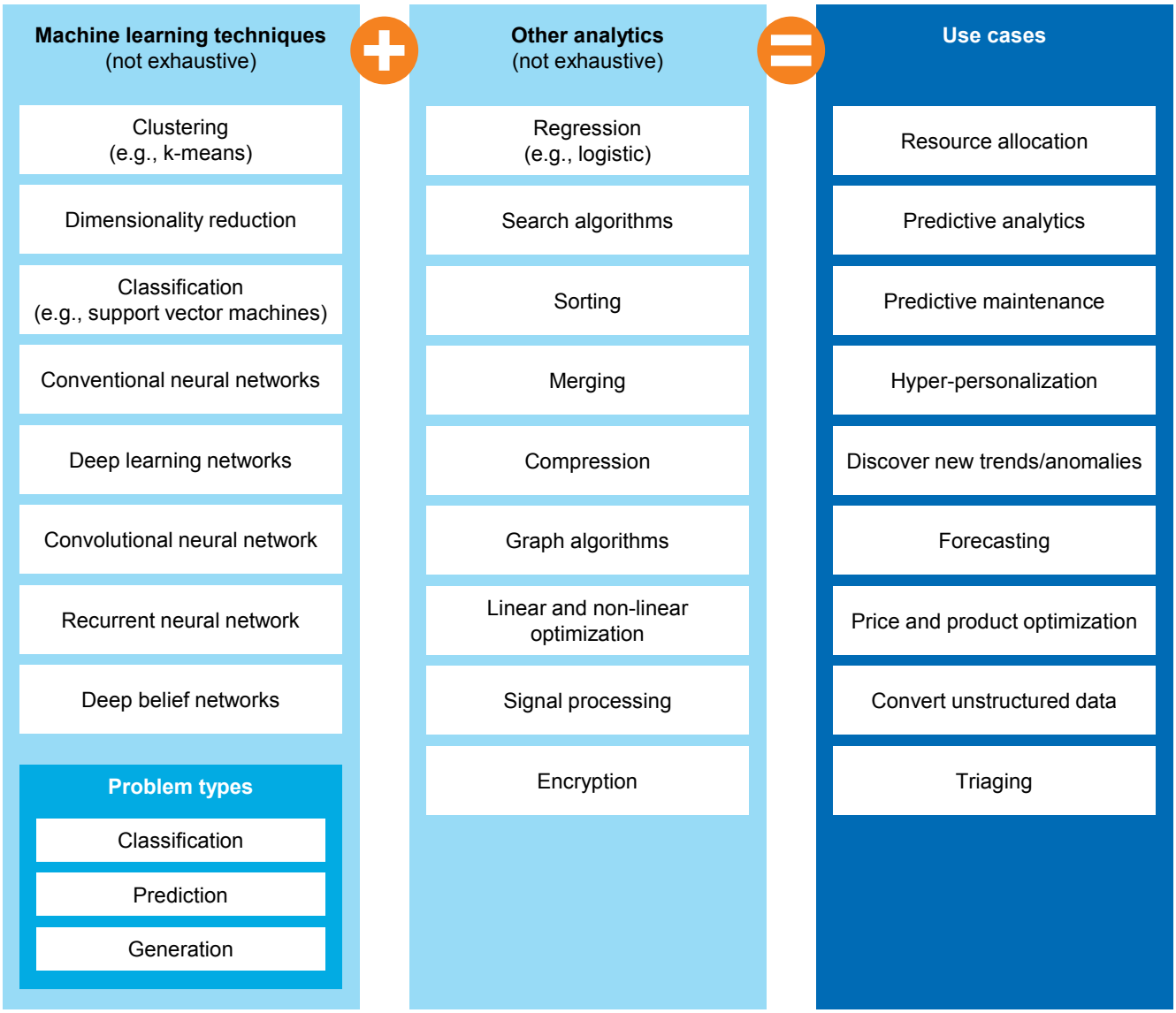
Machine learning can enhance the power of each of the archetypes described above. Conventional software programs are hard-coded by humans with specific instructions on the tasks they need to execute. By contrast, it is possible to create algorithms that "learn" from data without being explicitly programmed. The concept underpinning machine learning is to give the algorithm a massive number of "experiences" (training data) and a generalized strategy for learning, then let it identify patterns, associations, and insights from the data. In short, these systems are trained rather than programmed.

Some machine learning techniques, such as regressions, support vector machines, and k-means clustering, have been in use for decades. Others, while developed previously, have become viable only now that vast quantities of data and unprecedented processing power are available. Deep learning, a frontier area of research within machine learning, uses neural networks with many layers (hence the label "deep") to push the boundaries of machine capabilities. Data scientists have recently made breakthroughs using deep learning to recognize objects and faces and to understand and generate language. Reinforcement learning is used to identify the best actions to take now in order to reach some future goal. These type of problems are common in games but can be useful for solving dynamic optimization and control theory problems—exactly the type of issues that come up in modeling complex systems in fields such as engineering and economics. Transfer learning focuses on storing knowledge gained while solving one problem and applying it to a different problem. Machine learning, combined with other techniques, could have an enormous range of uses (see Exhibit E4 and Box E1, "The impact of machine learning").

This research offers a broad initial exploration of machine learning through two lenses. First, we investigate which business uses across 12 industries could be met by machine learning. Second, we examine which work activities currently performed by people could potentially be automated through machine learning and how that could play out across occupations. The initial findings here are meant to set the stage for future research.

Exhibit E4

Machine learning can be combined with other types of analytics to solve a large swath of business problems



SOURCE: McKinsey Global Institute analysis

Understanding the capabilities of machine learning and deep learning

Machine learning capabilities are best suited for solving three broad categories of problems: classification, prediction/estimation, and generation (Exhibit E5). Classification problems are about observing the world, including identifying objects in images and video, and recognizing text and audio. Classification also involves finding associations in data or segmenting it into clusters, which is useful in tasks such as customer segmentation. Machine learning can also be used to predict the likelihood of events and forecast outcomes. Lastly, it can be used to generate content, from interpolating missing data to generating the next frame in a video sequence.

Exhibit E5

Machine learning can help solve classification, prediction, and generation problems

Classification	Classify/label visual objects	Identify objects, faces in images and video
	Classify/label writing and text	Identify letters, symbols, words in writing sample
	Classify/label audio	Classify and label songs from audio samples
	Cluster, group other data	Segment objects (e.g., customers, product features) into categories, clusters
	Discover associations	Identify that people who watch certain TV shows also read certain books
Prediction	Predict probability of outcomes	Predict the probability that a customer will choose another provider
	Forecast	Trained on historical data, forecast demand for a product
	Value function estimation	Trained on thousands of games played, predict/estimate rewards from actions from future states for dynamic games
Generation	Generate visual objects	Trained on a set of artist's paintings, generate a new painting in the same style
	Generate writing and text	Trained on a historical text, fill in missing parts of a single page
	Generate audio	Generate a new potential recording in the same style/genre
	Generate other data	Trained on certain countries' weather data, fill in missing data points for countries with low data quality

SOURCE: McKinsey Global Institute analysis

Box E1. The impact of machine learning

Machine learning can be applied to a tremendous variety of problems—from keeping race cars running at peak performance to ferreting out fraud.

Off the track, Formula One (F1) teams compete in an arms race to make their cars faster. Top F1 teams pour hundreds of millions of dollars annually into development, continually aiming for incremental technological improvements that can boost speed. With so much at stake, F1 engineering teams constantly seek to improve productivity. Three F1 teams recently turned to machine learning to hold down costs in their aerodynamics operations divisions, which typically eat up more than 80 percent of development resources. Building on years of diverse project data—including CAD logs, human resources data, and employee communications—they looked for patterns that influenced the efficiency of an individual project. They discovered, for example, that too many engineers or long stoppages typically increased labor hours on a given project by 5 to 6 percent, while team use of the documentation system improved productivity by more than 4 percent. Overall, this application reduced the budget by 12 to 18 percent, saving millions of dollars.

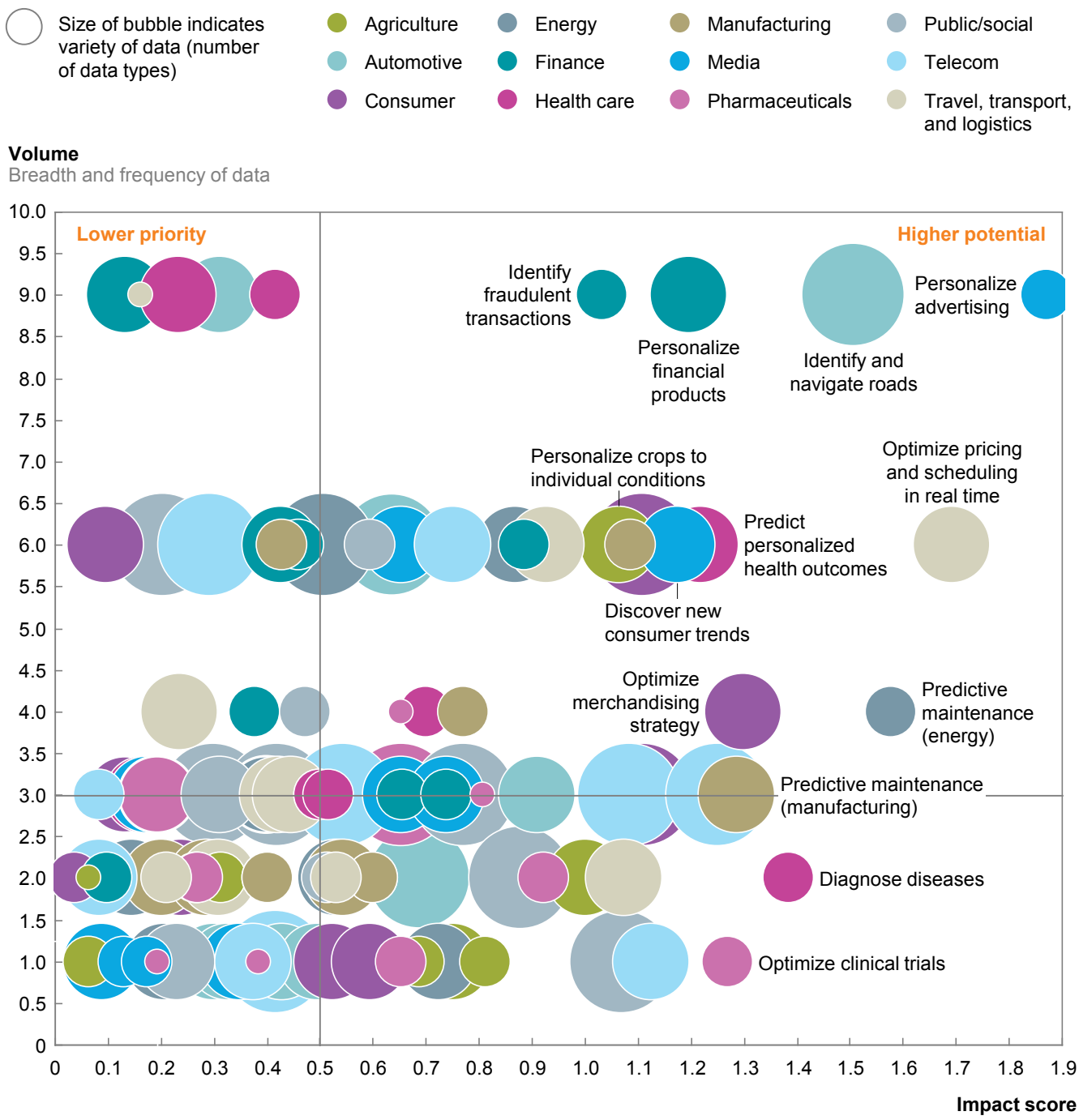
Another application of machine learning, predictive analytics, has proven to be effective at spotting fraud. At one large auto insurer, high accident rates for new client policies suggested that claims were being filed for pre-existing damage. The machine learning model was able to use diverse data to identify groups of new policies with accident rates six times those of the median. This grouping formed the basis of a new pricing strategy that improved profitability by more than 10 percent. Separately, a large retail bank in the United Kingdom used machine learning algorithms to identify fraudulent transactions with more than 90 percent accuracy. In another example, a large payment processor deployed machine learning on its extensive transaction data to identify "mule accounts" involved in money laundering.

We identified 120 potential use cases of machine learning in 12 industries, and surveyed more than 600 industry experts on their potential impact. The most striking finding was the extraordinary breadth of the potential applications of machine learning; each of the use cases was identified as being one of the top three in an industry by at least one expert in that industry. But there were differences.

We plotted the top 120 use cases in Exhibit E6. The y-axis shows the volume of available data (encompassing its breadth and frequency), while the x-axis shows the potential impact, based on surveys of more than 600 industry experts. The size of the bubble reflects the diversity of the available data sources.

Exhibit E6

Machine learning has broad potential across industries and use cases



SOURCE: McKinsey Global Institute analysis

The industry-specific uses that combine data richness with a larger opportunity are the largest bubbles in the top right quadrant of the chart. These represent areas where organizations should prioritize the use of machine learning and prepare for a transformation to take place. Some of the highest-opportunity use cases include personalized advertising; autonomous vehicles; optimizing pricing, routing, and scheduling based on real-time data in travel and logistics; predicting personalized health outcomes; and optimizing merchandising strategy in retail.

The use cases in the top right quadrant fall into four main categories. First is the radical personalization of products and services for customers in sectors such as consumer packaged goods, finance and insurance, health care, and media—an opportunity that most companies have yet to fully exploit. The second is predictive analytics. This includes examples such as triaging customer service calls; segmenting customers based on risk, churn, and purchasing patterns; identifying fraud and anomalies in banking and cybersecurity; and diagnosing diseases from scans, biopsies, and other data. The third category is strategic optimization, which includes uses such as merchandising and shelf optimization in retail, scheduling and assigning frontline workers, and optimizing teams and other resources across geographies and accounts. The fourth category is optimizing operations and logistics in real time, which includes automating plants and machinery to reduce errors and improve efficiency, and optimizing supply chain management.

Advances in deep learning could greatly expand the scope of automation

Previous MGI research examined the potential to automate 2,000 work activities performed in every occupation in the economy.[10] For each work activity, we identified the required level of machine performance across 18 human capabilities that could potentially enable automation.

Machine learning is particularly well-suited to implement seven of those 18 capabilities (Exhibit E7). The first striking observation is that almost all activities require some capabilities that correlate with what machine learning can do. In fact, only four out of more than 2,000 detailed work activities (or 0.2 percent) do not require any of the seven machine learning capabilities. Recognizing known patterns, by itself, is needed in 99 percent of all activities to varying degrees. This is not to say that such a high share of jobs is likely to be automated, but it does underscore the wide applicability of machine learning in many workplaces.

$3T
wages potentially affected if machine learning gains better capabilities in natural language understanding

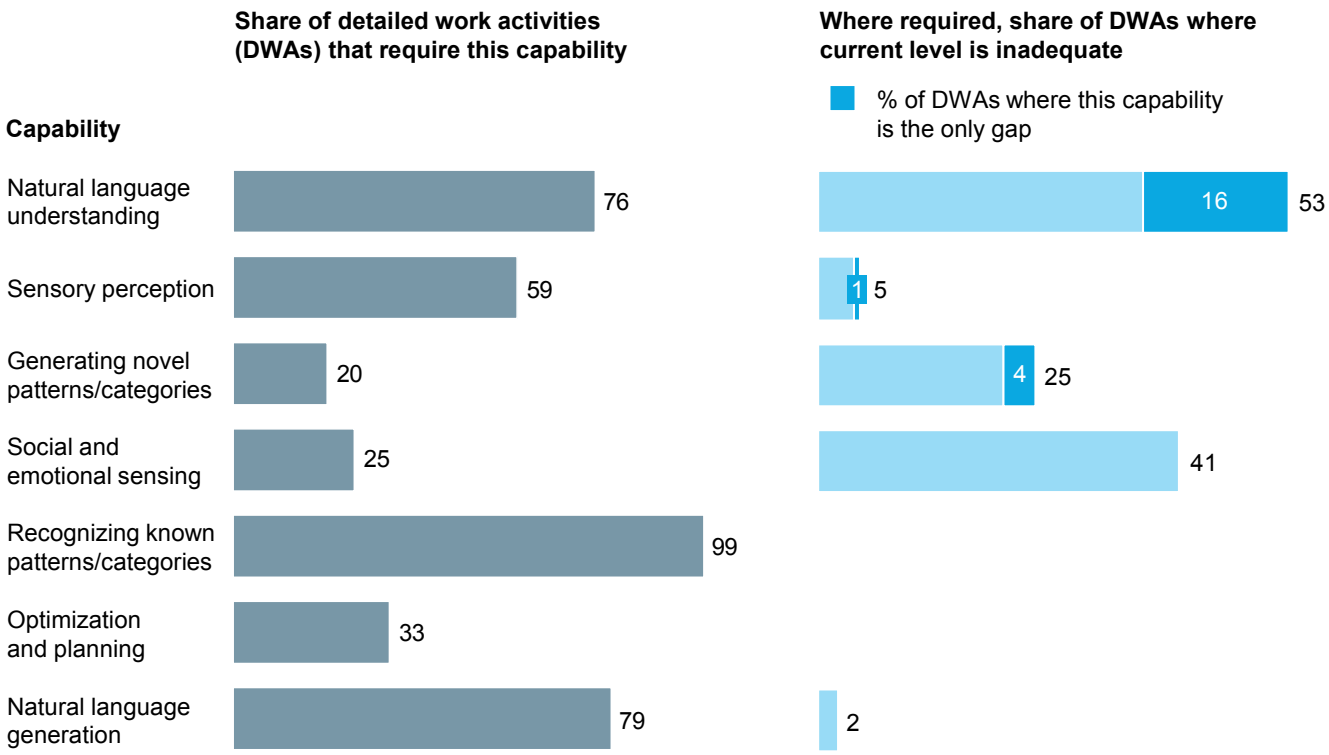
MGI's previous research on automation found that 45 percent of all work activities, associated with $14.6 trillion of wages globally, have the potential to be automated by adapting currently demonstrated technology. Some 80 percent of that could be implemented by using existing machine learning capabilities. But deep learning is in its early stages. Improvements in its capabilities, particularly in natural language understanding, suggest the potential for an even greater degree of automation. In 16 percent of work activities that require the use of language, for example, increasing the performance of machine learning in natural language understanding is the only barrier to automation. Improving natural language capabilities alone could lead to an additional $3 trillion in potential global wage impact.

[10] These "detailed work activities" are defined by O*NET, a data collection program sponsored by the US Department of Labor. See Michael Chui, James Manyika, and Mehdi Miremadi, "Four fundamentals of workplace automation," *McKinsey Quarterly*, November 2015.

Exhibit E7

Improvements in natural learning understanding and generation as well as social sensing would have the biggest impact on expanding the number of work activities that deep learning could technically automate

%



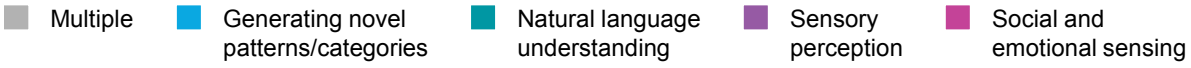
SOURCE: McKinsey Global Institute analysis

We further looked at which occupations that could be affected by improvements in deep learning represent the greatest potential wage impact (Exhibit E8). The role of customer service representatives, in particular, lends itself to automation across most of its work activities. Deep learning is also likely to have a large impact on frontline supervisory roles and in occupations with primarily administrative duties, including executive assistants, cashiers, and waitstaff. Large numbers of people are employed in these occupations, which points to the possibility of substantial job displacement. In addition, advances in machine learning could automate significant percentages of the activities associated with some high-paying jobs such as lawyers and nurses.

While machine learning in general and deep learning in particular have exciting and wide-ranging potential, there are real concerns associated with their development and potential deployment. Some of these, such as privacy, data security, and data ownership, were present even before the big data age. But today new questions have formed.

Exhibit E8

Improvements in deep learning (DL) could affect billions of dollars in wages in ten occupations globally



Occupations	% of time spent on activities that could be automated if DL improves (by DWA group)[1]	Most frequently performed group of DWAs that could be automated if DL improves	Global employ-ment Million	Hourly wage $	Global wages that DL could automate $ billion
Secretaries and administrative assistants, except legal, medical, and executive	28	Interacting with computers to enter data, process information, etc.	48.2	3.90	109.8
Business operations specialists, all other	30	Performing administrative activities	6.1	24.68	94.2
Managers, all other	27	Monitoring processes, materials, or surroundings	8.3	18.25	86.7
First-line supervisors of office and administrative support workers	35	Interpreting the meaning of information for others	12.8	8.75	81.5
Cashiers	18	Performing administrative activities	68.1	3.18	81.5
First-line supervisors of retail sales workers	13	Guiding, directing, and motivating subordinates	19.7	15.02	77.4
Industrial engineers	20	Getting information	8.0	20.60	69.4
Customer service representatives	51	Performing for or working directly with the public	6.9	9.35	67.4
Lawyers	31	Providing consultation and advice to others	2.3	41.14	61.8
First-line supervisors of helpers, laborers, and material movers	24	Organizing, planning, and prioritizing work	8.5	12.73	54.2

1 Detailed work activity. There are 37 total DWA groups.

SOURCE: National labor and statistical sources; McKinsey Global Institute analysis

First, deep learning models are opaque, which can be a barrier to adoption in certain applications. As of today, it is difficult to decipher how deep neural networks reach insights and conclusions, making their use challenging in cases where transparency of decision making may be needed for regulatory purposes. Also, decision makers and customers may not buy into insights generated in a non-transparent way, especially when those insights are counterintuitive.

Second, there are ethical questions surrounding machine intelligence. One set of ethical concerns relates to real-world biases that might be embedded into training data. Another question involves deciding whose ethical guidelines will be encoded in the decision making of intelligence and who is responsible for the algorithm's conclusions. Leading artificial intelligence experts, through OpenAI, the Foundation for Responsible Robotics, and other efforts, have begun tackling these questions.

Third, the potential risks of labor disruption from the use of deep learning to automate activities are generating anxiety. There is historical precedent for major shifts among sectors and changes in the nature of jobs in previous waves of automation. In the United States, the share of farm employment fell from 40 percent in 1900 to 2 percent in 2000; similarly, the share of manufacturing employment fell from roughly 25 percent in 1950 to less than 10 percent in 2010. In both circumstances, while some jobs disappeared, new ones were created, although what those new jobs would be could not be ascertained at the time. But history does not necessarily provide assurance that sufficient numbers of new, quality jobs will be created at the right pace. At the same time, many countries have or will soon have labor forces that are declining in size, requiring an acceleration of productivity to maintain anticipated rates of economic growth. But automation technologies will not be widely adopted overnight; in fact, a forthcoming MGI research report will explore the potential pace of automation of different activities in different economies. Certainly dealing with job displacement, retraining, and unemployment will require a complex interplay of government, private sector, and educational and training institutions, and it will be a significant debate and an ongoing challenge across society.

•••

Data and analytics have even greater potential to create value today than they did when companies first began using them. Organizations that are able to harness these capabilities effectively will be able to create significant value and differentiate themselves, while others will find themselves increasingly at a disadvantage.

© Cultura Creative/Alamy

RELATED MGI AND MCKINSEY RESEARCH

Big Data: The next frontier for innovation, competition, and productivity (May 2011)

Big data will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus—as long as the right policies and enablers are in place.

The Internet of Things: Mapping the value beyond the hype (June 2015)

If policymakers and businesses get it right, linking the physical and digital worlds could generate up to $1.1 trillion a year in economic value by 2025.

Disruptive technologies: Advances that will transform life, business, and the global economy (May 2013)

Twelve emerging technologies—including the mobile Internet, autonomous vehicles, and advanced genomics—have the potential to truly reshape the world in which we live and work. Leaders in both government and business must not only know what's on the horizon but also start preparing for its impact.

Digital America: A tale of the haves and have-mores (December 2015)

While the most advanced sectors, companies, and individuals push the boundaries of technology use, the US economy as a whole is realizing only 18 percent of its digital potential.

www.mckinsey.com/mgi

E-book versions of selected MGI reports are available at MGI's website, Amazon's Kindle bookstore, and Apple's iBooks Store.

Download and listen to MGI podcasts on iTunes or at www.mckinsey.com/mgi/publications/multimedia/

Cover image: © Nadla/Getty Images.

McKinsey Global Institute
December 2016
Copyright © McKinsey & Company
www.mckinsey.com/mgi

@McKinsey_MGI
McKinseyGlobalInstitute